

Effective Vaccination Policies

L. Shaw^a, W. Spears^{*,b}, L. Billings^c, P. Maxim^d

^a*Department of Computer Science, University of Wyoming, Laramie, WY 82070 USA*

^b*Swarmotics LLC, Laramie, WY 82070 USA*

^c*Department of Mathematical Sciences, Montclair State University, Montclair, NJ 07043 USA*

^d*Wyoming Department of Transportation, Cheyenne, WY, 82009, USA*

Abstract

We present a framework for modeling the spread of pathogens throughout a population and generating policies that minimize the impact of those pathogens on the population. This framework is used to study the spread of human viruses between cities via airplane travel. It combines agent-based simulation, mathematical analysis, and an Evolutionary Algorithm (EA) optimizer. The goal of this study is to develop tools that determine the optimal distribution of a vaccine supply in the model. Using plausible benchmark vaccine allocation policies of uniform and proportional distribution, we compared their effectiveness to policies found by the EA. We then designed and tested a new, more effective policy which increased the importance of vaccinating smaller cities that are flown to more often. This “importance factor” was validated using U.S. influenza data from the last four years.

Key words: epidemiology, evolutionary algorithm, influenza, migration, vaccination.

1. Introduction

At the beginning of each influenza season, there are widespread concerns in the U.S. about the supply and distribution of the latest vaccine. Public fear is fuelled by stories of ineffective influenza vaccine batches, a limited availability of the vaccine, and deadly new strains that could possibly become pandemic. Identifying the urgent need for an effective vaccination policy, this paper presents a novel framework for modeling the spread of pathogens throughout a population

*Corresponding author

Email address: wspears@swarmotics.com (W. Spears)

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 31 MAR 2010		2. REPORT TYPE		3. DATES COVERED 00-00-2010 to 00-00-2010	
4. TITLE AND SUBTITLE Effective Vaccination Policies		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Wyoming, Department of Computer Science, Laramie, WY, 82070		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES Information Sciences, 180 (19), 3728?3744					
14. ABSTRACT We present a framework for modeling the spread of pathogens throughout a population and generating policies that minimize the impact of those pathogens on the population. This framework is used to study the spread of human viruses between cities via airplane travel. It combines agent-based simulation, mathematical analysis, and an Evolutionary Algorithm (EA) optimizer. The goal of this study is to develop tools that determine the optimal distribution of a vaccine supply in the model. Using plausible benchmark vaccine allocation policies of uniform and proportional distribution, we compared their effectiveness to policies found by the EA. We then designed and tested a new, more effective policy which increased the importance of vaccinating smaller cities that are flown to more often. This ?importance factor? was validated using U.S. influenza data from the last four years.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 31	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

and a way to generate and evaluate policies that minimize the impact of those pathogens on the population. This framework is a combination of agent-based simulation, mathematical analysis, and a sophisticated optimization tool.

The objective of our model is to study of the spread of human viruses between cities via air travel. Previous continuous models of the spread of disease between cities include Rvachev and Longini [26], Hyman and LaForce [16], Colizza, et al. [9], and Grais, et al. [11]. These types of epidemiological models use differential equations with mass action [1, 20, 25] or migration [2, 18] coupling terms between subpopulations. Another approach is in agent-based simulations of disease spread, such as [21, 10]. While these models fold the transport between weakly connected subpopulations into the existing model, we layer the model into two parts. We use an agent-based simulation to model the movement between the subpopulations and differential equations to govern the state of the agents within each subpopulation. The complete simulation is parallelized by running each island (subpopulation) on a separate computer processor.

We extend the model to include a supply of vaccine, which is to be distributed in an optimal fashion. By mathematical analysis of this model, we determine the feasibility of a vaccination policy. Using a sophisticated in-house optimization toolkit [27, 28], based on Evolutionary Algorithms (EA), we search for vaccination policies that minimize the number of infected people. We compare these policies to plausible benchmark policies to verify that the EA policies are more effective. Analysis of the EA policies indicates that the vaccine is generally distributed to (1) the city of origin for the virus, (2) cities that are traveled to most often, and (3) smaller cities. The latter observation is most surprising and counter-intuitive.

Finally, we design a new benchmark policy to improve upon the EA results, based on the three observations given above. This new policy is superior to all of our benchmark policies and the EA-generated policies. The key to the performance is an “importance factor”, which emphasizes the vaccination of small cities that are flown to most often. Confirmation of this “importance factor” is provided by real-world data that monitored the spread of the influenza virus during the winters of the last four years. We believe this new policy is quite general. However, if the model changes sufficiently, new policies will eventually be required. In that case the EA is always available to guide the human in the creation of the new policies for changing situations.

2. The Model

Our virus spread model has two basic levels – an inter-city level, called the “micro-level”, and the intra-city level, called the “macro-level” or air-travel model. The micro-level model is governed by differential equations (the island part of the model, e.g., see [4, 30]). Connecting the islands is the macro-level model, using a Markov chain probability transition matrix, that simulates the intra-city traffic by airline flights (the weakly-connected part of the model). The combination of levels allows us to model roughly 55 million people within the United States.

2.1. The Micro-Level Model

To simulate the disease spread within a city, we use a compartmental model that assumes the population transitions between Susceptible, Asymptomatic, Infected, and Recovered states, called the SAIR model. The Susceptible, Infected, and Recovered states are standard states of health used in SIR models. We also use an Asymptomatic state, which represents the period of time after initial infection that a person is unaware of the infection (due to a complete lack of symptoms) but is contagious [13]. We also assume that a person in that state may fly in the macro-level model, while an infected person can not fly.

The variables that represent the proportion of the city’s population in each state are: the susceptible state s , the asymptomatic state a , the infected state i , and the recovered state r . The proportions are changed each time-step of the simulation according to the following system of differential equations:

$$\begin{aligned} \frac{ds}{dt} &= \delta' r - \alpha s \\ \frac{da}{dt} &= \alpha s - \mu a \\ \frac{di}{dt} &= \mu a - \delta i \\ \frac{dr}{dt} &= \delta i - \delta' r. \end{aligned} \tag{1}$$

The parameters α , μ , δ , and δ' are the probabilities of an individual transitioning between states, on a daily basis. We can think of α as the infection rate, $1/\mu$ as the expected asymptomatic time of the virus (how long an individual is contagious before showing symptoms), $1/\delta$ as the expected length of time a person spends infected until recovery, and $1/\delta'$ as the expected time it takes for an individual to lose resistance to the virus. The population in the absence of air travel is assumed to be constant and we have the relationship $s + a + i + r = 1$. The parameters μ , δ , and δ' are constants. Figure 1 gives a diagram of the health states and the transitions between them.

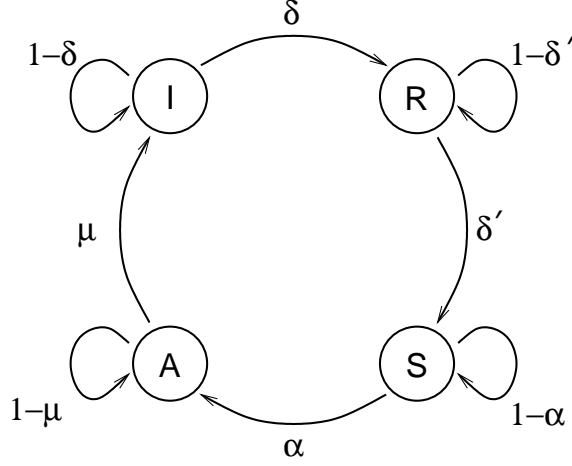


Figure 1: SAIR State Diagram

The infection rate (α) is related to the proportion of the population that can infect an individual (i.e. a and i). The probability of an individual being infected is related to the number of people the individual interacts with (\bar{b}) and the probability the individual will be infected from an encounter with an infected person (β). Since β is the probability of being infected by an infected neighbor, $1 - \beta$ is the probability of *not* being infected by an infected neighbor. We know $a + i$ is the proportion of the population that is contagious and \bar{b} is the number of neighbors a person has, so there are approximately $(a + i)\bar{b}$ infected neighbors for a particular person. This yields a probability of not being infected of $(1 - \beta)^{(a+i)\bar{b}}$. But since we are interested in the probability of being infected, we use the complement of the probability of not being infected for α . This yields the following equation which can be interpreted to be the probability an individual will be infected by at least one of his/her \bar{b} neighbors:

$$\alpha = 1 - (1 - \beta)^{(a+i)\bar{b}}. \quad (2)$$

In the simulation, we use this expression to compute the infection rate.

The constant parameters μ , δ , and δ' are assumed to be the same for each city. The infection rate (α) is the only variable that could change between cities, because it depends on \bar{b} (which is related to the population density). Our airports are in large U.S. cities, so there shouldn't be significant changes in \bar{b} . Further, the ODE model is not sensitive to small changes in α (the dynamics are qualitatively similar) and the infection spread will just slightly speed up or slow down. Therefore, small changes in that parameter should not improve the accuracy drastically.

We continue with a brief overview of the steady state solutions of this system and their stabilities. These fixed points will be written in the form (s, a, i, r) . For the analysis of the behavior of the differential equations, we use the following approximation for small β and $(a + i)\bar{b}$,

$$\alpha \approx 1 - (1 - \beta\bar{b}(a + i)) = \beta\bar{b}(a + i). \quad (3)$$

The disease free equilibrium (DFE), which represents the die out of the disease, is the point $(1, 0, 0, 0)$. A common method of determining its local stability is linearizing the system about the steady state. The resulting eigenvalues are

$$\begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= -\delta' \\ \lambda_3 &= \frac{\beta\bar{b} - (\delta + \mu) + \sqrt{(\beta\bar{b} - (\delta + \mu))^2 + 4(\beta\bar{b}(\delta + \mu) - \mu\delta)}}{2} \\ \lambda_4 &= \frac{\beta\bar{b} - (\delta + \mu) - \sqrt{(\beta\bar{b} - (\delta + \mu))^2 + 4(\beta\bar{b}(\delta + \mu) - \mu\delta)}}{2}. \end{aligned} \quad (4)$$

We have local stability when these eigenvalues are nonpositive, or $\frac{\beta\bar{b}(\mu + \delta)}{\mu\delta} < 1$.

The other steady state solution represents the endemic equilibrium (s_e, a_e, i_e, r_e) , which is given by

$$\begin{aligned} s_e &= \frac{\mu\delta}{\beta\bar{b}(\delta + \mu)} \\ a_e &= \frac{1 - s_e}{1 + \frac{\mu}{\delta} + \frac{\mu}{\delta'}} \\ i_e &= \frac{\mu a_e}{\delta} \\ r_e &= \frac{\mu a_e}{\delta'}. \end{aligned} \quad (5)$$

To find the replacement number R , we follow Hethcote, et al. [12]. Assuming that a person is contagious in the asymptomatic and infected states, a person is infective for about $1/\mu + 1/\delta$ days. If a person is contagious, he/she will infect approximately $\beta\bar{b}$ people, so we can estimate R as follows:

$$R = \frac{\beta\bar{b}(\delta + \mu)}{\mu\delta} = \frac{1}{s_e}. \quad (6)$$

When $R < 1$, the virus dies off. This agrees with the local stability analysis of the DFE. When $R > 1$, the virus spreads.

2.2. The Macro-Level Model

The micro-level model for the cities provides the island part of the weakly connected island modeling framework, but we need to provide the connections

between the islands. We are interested in the effect of airplane travel between cities as a means of spreading a virus. To model this, we searched for data on airplane travel in the United States to figure out which cities have the biggest airports. As a first-order approximation, we concentrated on the the top 12 U.S. airports. Table 1 shows the 12 cities chosen ordered by population size with each city's yearly total passengers.

<i>City</i>	<i>Pop. Size</i>	<i>Passengers/Year</i>	<i>% of Total</i>
Las Vegas	1.314	35.009	6.742
Denver	1.985	35.651	6.866
Minneapolis	2.389	32.628	6.284
Phoenix	2.907	35.547	6.846
Atlanta	3.500	76.876	14.806
Houston	3.823	33.905	6.530
Detroit	3.903	32.477	6.255
Dallas	4.146	52.828	10.174
San Francisco	4.767	31.456	6.058
Miami	4.919	30.060	5.789
Chicago	8.307	66.565	12.820
Los Angeles	13.296	56.223	10.828
Total	55.256	519.229	100.000

Table 1: Cities Ordered by Population Size (in millions)

We used this information to estimate the percentage of passengers for each city, which yielded a initial probability distribution for allocating planes to cities. We then computed the frequency of flights between the cities. We booked non-stop flights between each city on a particular Wednesday to estimate the number of flights between each city on <http://www.expedia.com>. Since there are twelve cities, we booked 132 (11×12) different flights and counted the number of non-stop flights between the cities. Table 2 shows the number of flights between each city.

To convert the number of flights into transition probabilities, we divide each entry in the matrix in Table 2 by its row sum, yielding the probability Q_{ij} of traveling from the row city i to the column city j . Table 3 shows the resulting Q -matrix. Note that the table is not symmetric about the diagonal, so $Q_{ij} \neq Q_{ji}$. Q is ergodic and has an equilibrium distribution.

For the twelve cities, the number of daily flights is 1979, (see Table 2). According to the fleet composition of United Airlines we assume that each plane can carry between 160 and 300 passengers. Assuming that the average flight carries roughly 130 passengers (at least 43% to 82% of capacity), this yields about 95 million passengers per year. Since we are modeling roughly 18% of the airport

City	Atl	Chi	LA	Dal	Den	Pho	LV	Hou	Minn	Det	SF	Mia	Total
Atl	0	21	18	23	23	14	10	22	17	16	15	19	198
Chi	20	0	18	22	18	20	18	19	22	22	22	19	220
LA	17	18	0	21	18	21	16	15	12	7	17	14	176
Dal	21	22	20	0	19	18	19	18	17	14	21	9	198
Den	22	21	17	20	0	20	20	18	21	10	18	8	195
Pho	14	20	19	20	20	0	11	16	16	12	22	2	172
LV	11	17	16	19	20	11	0	7	8	7	18	3	137
Hou	23	18	15	16	18	16	15	0	9	9	13	8	160
Minn	18	22	11	17	20	16	8	9	0	15	8	3	147
Det	17	21	7	14	10	12	7	9	14	0	3	6	120
SF	12	22	18	20	19	22	17	13	8	3	0	6	160
Mia	20	17	14	9	9	2	3	8	3	6	5	0	96
Total	195	219	173	201	194	172	144	154	147	121	162	97	1979

Table 2: Number of Flights Between Each City

	Atl	Chi	LA	Dal	Den	Pho	LV	Hou	Minn	Det	SF	Mia
Atl	0.000	0.106	0.091	0.116	0.116	0.071	0.051	0.111	0.086	0.081	0.076	0.096
Chi	0.091	0.000	0.082	0.100	0.082	0.091	0.082	0.086	0.100	0.100	0.100	0.086
LA	0.097	0.102	0.000	0.119	0.102	0.119	0.091	0.085	0.068	0.040	0.097	0.080
Dal	0.106	0.111	0.101	0.000	0.096	0.091	0.096	0.091	0.086	0.071	0.106	0.046
Den	0.113	0.108	0.087	0.103	0.000	0.103	0.103	0.092	0.108	0.051	0.092	0.041
Pho	0.081	0.116	0.111	0.116	0.116	0.000	0.064	0.093	0.093	0.070	0.128	0.012
LV	0.080	0.124	0.117	0.139	0.146	0.080	0.000	0.051	0.058	0.051	0.131	0.022
Hou	0.144	0.113	0.094	0.100	0.113	0.100	0.094	0.000	0.056	0.056	0.081	0.050
Minn	0.122	0.150	0.075	0.116	0.136	0.109	0.054	0.061	0.000	0.102	0.054	0.020
Det	0.142	0.175	0.058	0.117	0.083	0.100	0.058	0.075	0.117	0.000	0.025	0.050
SF	0.075	0.138	0.113	0.125	0.119	0.138	0.106	0.081	0.050	0.019	0.000	0.038
Mia	0.208	0.177	0.146	0.094	0.094	0.021	0.031	0.083	0.031	0.063	0.052	0.000
Total	1.259	1.419	1.073	1.244	1.203	1.022	0.830	0.911	0.853	0.703	0.943	0.540

Table 3: Probability Transition Matrix Q

traffic, this is consistent with the estimate of 519 million passengers per year (see Table 1).

The connection between the macro-level and micro-level models is as follows. We assume a well-mixed population and we randomly sample a group to migrate from city to city. Homogeneity inside city populations is a standard assumption in these models (e.g., see [8]). Over many trials, the average of the randomly picked group should represent the properties of the population. Specifically, the arrivals deplane, and then a random sample of the population gets on the plane. Hence, at the beginning of each flight, the proportions of susceptible, asymptomatic, and recovered (but not infected) people on the plane are a sample of the population of the city the plane currently resides in. Infected people are assumed to be too sick to fly. We also do not model the spread of the disease within the airplane itself.

3. Introducing Vaccines

Thus far our model has not included any mechanisms for containment. There are many types of containment, including vaccines, anti-viral medications, and quarantine. In this paper we focus on vaccines, and extend our mathematical model to include vaccinations.

We model vaccinated individuals as people with a decreased probability of being infected by an infected neighbor. So we effectively have two subpopulations in our model - vaccinated and unvaccinated people. We make a simplifying assumption that vaccination has no effect on the asymptomatic and infected times. With that simplifying assumption, we only need to change the infection rate α . Let f be the proportion of the population that is vaccinated ($0 \leq f \leq 1$). We assume f applies to each health state equally - that is f is the proportion of the susceptible people that are vaccinated as well as the proportion of the asymptomatic people that have been vaccinated, etc. Let β be the probability an unvaccinated person is infected by an infected neighbor and $\beta' < \beta$ be the probability a vaccinated person is infected by an infected neighbor. Therefore, we have two infection rates - one for the unvaccinated (α) and the other for the vaccinated (α') proportions of the population,

$$\begin{aligned}\alpha &= 1 - (1 - \beta)^{(a+i)\bar{b}} \\ \alpha' &= 1 - (1 - \beta')^{(a+i)\bar{b}}.\end{aligned}\tag{7}$$

The average probability of infection $\bar{\alpha}$ is given by

$$\bar{\alpha} = f\alpha' + (1 - f)\alpha.\tag{8}$$

As before, we can approximate $\bar{\alpha}$ as

$$\bar{\alpha} \approx (a + i)\bar{b}(f\beta' + (1 - f)\beta).\tag{9}$$

Hence, the new differential equations model with vaccination is

$$\begin{aligned}\frac{ds}{dt} &= \delta'r - \bar{\alpha}s \\ \frac{da}{dt} &= \bar{\alpha}s - \mu a \\ \frac{di}{dt} &= \mu a - \delta i \\ \frac{dr}{dt} &= \delta i - \delta'r.\end{aligned}\tag{10}$$

The disease free equilibrium (DFE) is the point $(1, 0, 0, 0)$. The eigenvalues from the linearization about this point is

$$\begin{aligned}\lambda_1 &= 0 \\ \lambda_2 &= -\delta' \\ \lambda_3 &= \frac{\bar{b}\omega - (\delta + \mu) + \sqrt{(\bar{b}\omega - (\delta + \mu))^2 + 4(\bar{b}\omega(\delta + \mu) - \mu\delta)}}{2} \\ \lambda_4 &= \frac{\bar{b}\omega - (\delta + \mu) - \sqrt{(\bar{b}\omega - (\delta + \mu))^2 + 4(\bar{b}\omega(\delta + \mu) - \mu\delta)}}{2},\end{aligned}\tag{11}$$

with $\omega = (f\beta' + (1-f)\beta)$. The DFE has local stability when $\frac{\bar{b}(f\beta' + (1-f)\beta)(\mu + \delta)}{\mu\delta} < 1$.

The other steady state solution represents the endemic equilibrium, which is given by

$$\begin{aligned}s_e &= \frac{\delta\mu}{\bar{b}(\delta + \mu)(f\beta' + (1-f)\beta)} \\ a_e &= \frac{1 - s_e}{1 + \frac{\mu}{\delta} + \frac{\mu}{\delta'}} \\ i_e &= \frac{a_e\mu}{\delta} \\ r_e &= \frac{a_e\mu}{\delta'}.\end{aligned}$$

Then, the replacement number R is

$$R = \bar{b}(f\beta' + (1-f)\beta)\left(\frac{1}{\mu} + \frac{1}{\delta}\right).\tag{12}$$

As before, when $R < 1$, the virus dies off. When $R > 1$, the virus spreads.

It is helpful to quantify the lower limit for an effective vaccination rate f_e . When $R = 1$,

$$f_e = \frac{\frac{\delta\mu}{\bar{b}(\delta + \mu)} - \beta}{\beta' - \beta}.\tag{13}$$

We can define the “vaccine feasibility region” as follows:

$$0 \leq f_e < f \leq 1.\tag{14}$$

If $f_e < 1$, then it is feasible to stop the spread of the virus. Accomplishing this requires that a fraction $f_e < f \leq 1$ of the population be vaccinated. If $f < f_e$ or $f_e > 1$, then the vaccine merely slows the spread of the virus. An “effective” vaccine can stop the virus spread by causing herd immunity and the disease will die out. Vaccines that have this capability are in the vaccine feasibility region. Vaccines not in this region will only lower the endemic threshold of the population (the stable state of infected individuals).

Figure 2 shows an example of a theoretically derived feasibility region for a reasonable choice of parameter settings. The parameter settings are inspired by a typical influenza virus. With $\mu = 0.35$, the asymptomatic time is almost 3 days. The period of infection is roughly one week ($\delta = 0.15$). Finally, we assume that it takes roughly three years before you become susceptible again ($\delta' = 0.001$). The upper left area above the curve in black is the region of feasible vaccine efficacy, where the efficacy is measured by β' (the probability an infective will infect a vaccinated susceptible). Note that even if the vaccine is perfect ($\beta' = 0$), then greater than 62% of the population must be vaccinated to eliminate the virus spread. As the vaccine efficacy decreases (higher β'), then that fraction increases. Once $\beta' > 0.012$, vaccines will merely slow the spread of the virus.

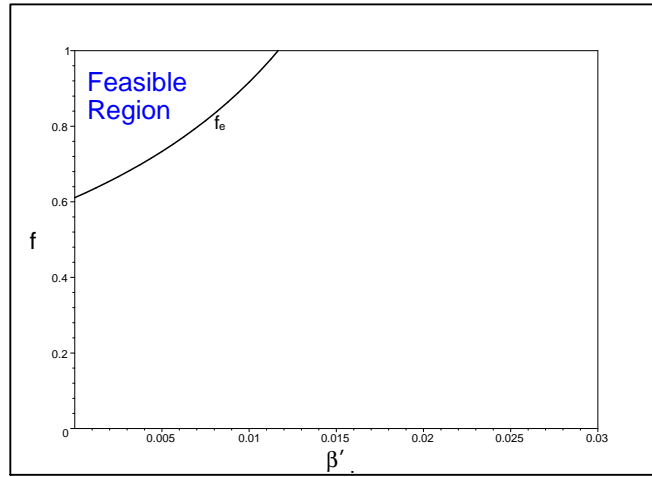


Figure 2: Theoretical Vaccine Feasibility Region ($\bar{b} = 9$, $\mu = 0.35$, $\beta = 0.03$, $\delta = 0.15$, $\delta' = 0.001$)

We compared the theoretical prediction with our simulation. We ran our simulation for 300 days (using the same parameter settings), and monitored the total number of “sick days” (we do not model births or deaths in our simulation). Figure 3 summarizes the results of these experiments. This graph was generated by running the simulation with $f = 0.0$ to $f = 1.0$ in increments of 0.01 and with $\beta' = 0$ to $\beta' = \beta = 0.03$ in increments of 0.001. The colors represent the sum of sick days taken after 300 days of simulation time (i.e. the sum of the infected population after each day for 300 days). Inside the feasibility region the virus resulted in less than one million sick days. Outside of that region the number of sick days increased enormously. The boundary of the region represents a phase transition.

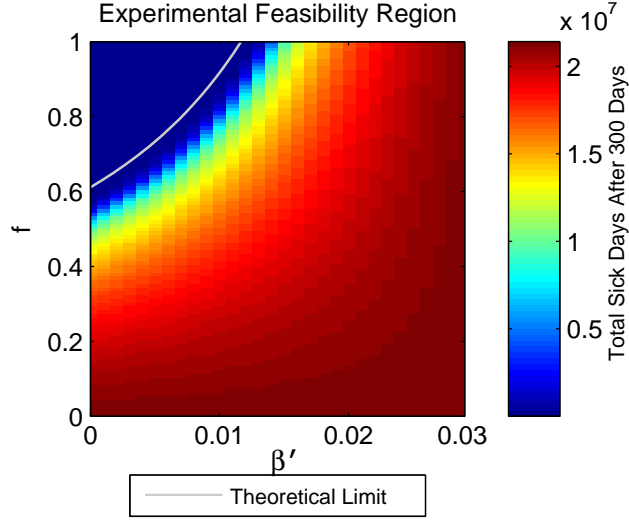


Figure 3: Experimental Vaccine Feasibility Region ($\bar{b} = 9$, $\beta = 0.03$, $\mu = 0.35$, $\delta = 0.15$, $\delta = 0.001$)

As a brief implementation note, our simulation is written in C++, using MPI to allow for the use of multiple processors. The simulation was thoroughly verified and validated to ensure agreement with the differential equation steady-state distribution and the Markov chain equilibrium distribution.¹ We used a Beowulf cluster with 13 processors. One processor was used for each of the 12 cities. A thirteenth processor was used to control the communication (the flights) between each city. One year of simulation time takes less than one half a minute of real time.

4. Benchmark Vaccination Policies

In order to better understand the effects of vaccination on our model, we created several benchmark vaccination policies and examined their performance over several scenarios. In each scenario the virus starts as a small population of 100 asymptomatic people in one city. In order to examine the effects of city population size, we chose three cities – Las Vegas, Atlanta and Los Angeles. Las Vegas

¹In addition, a separate StarLogo implementation was written, and all results were within 1% of the C++ version

is the smallest city, while Los Angeles is the largest. Atlanta is a city of average size.

For each city we also examined three vaccination delays (from the day the infection starts) of 15, 30, and 45 days. These delays can be considered to be a sum of two effects, (1) a delay in giving vaccination shots, and (2) a delay in how long it takes for the vaccine to take effect.

Using the influenza virus parameters mentioned earlier in the paper, the proportion of the population that must be vaccinated to ensure that the virus is quickly eliminated is $f > f_e \approx 92\%$ (assuming $\beta' = 0.01$). Since the total population size is roughly 55 million people, theory indicates that at least roughly 51 million vaccines are required to quickly eliminate the virus. However, this is unlikely to be the case, so we consider having the number of vaccines V range from 5 million to 55 million, in increments of 5 million. Hence, when there are less than 50 million vaccines, it is impossible to completely stop the spread of the virus. Instead, our goal is to allocate the vaccines in such a way as to minimize the impact. Our measure of “impact” is the total number of sick days taken by the population over 450 days.

We first considered two possible benchmark vaccination policies. The “uniform” vaccination strategy allocates $1/N$ of the number of vaccines to each of the N cities. In this model, $N = 12$. The “proportional” vaccination strategy allocates vaccines according to population size (e.g., a city that has twice as many people will receive twice as many vaccines). We also considered a possible modification to these two policies. Preliminary runs indicated that fully vaccinating the city of viral origin was important. This turned out to always hold, so our policies were modified as follows:

<i>Name</i>	Description
Uniform	Fully vaccinate the city of origin, uniformly distributing the rest to the remaining cities
Proportional	Fully vaccinate the city of origin, distributing the rest to the remaining cities according to their proportion total population minus the population of the city of origin.

Figure 4: Two Benchmark Vaccine Allocation Policies

Our results were remarkably consistent, regardless of the city of origin and the vaccine delay. Hence, in this paper we present only the results for Atlanta with a 30 day delay in vaccination. In Figure 5, we see that the proportional and uniform allocation policies are roughly equivalent up to 20 million vaccines.² Between 25 and 40 million vaccines, the uniform allocation policy was considerably better than the proportional policy. Above 40 million vaccines, the two policies are roughly equivalent again. Overall, the uniform allocation of vaccines was definitely superior.

This result is counter-intuitive. There is a two-fold explanation. First, the probability of infection depends on the *proportion* of asymptomatic and infected people. Hence, having the virus start as a small population of 100 asymptomatic people yields a smaller proportion in larger cities. One can consider this to be a “dilution factor” and explains why outbreaks appear to spread much faster in very small communities. Second, the passengers on a plane are a representative sampling of the city that the plane is in. The odds of an asymptomatic person boarding the plane are much higher in a small city. Hence, it is actually better to vaccinate smaller cities.

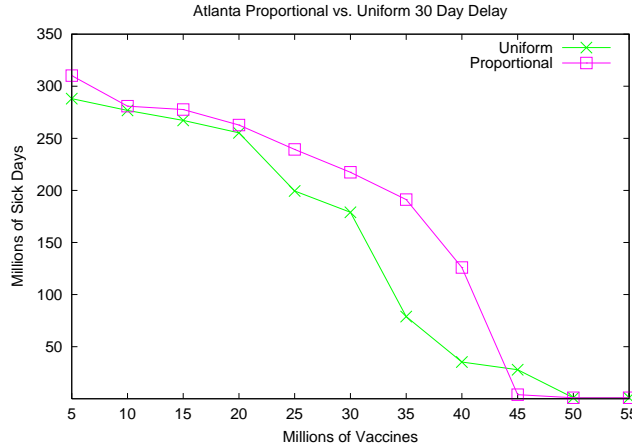


Figure 5: Atlanta Uniform and Proportional Baselines Comparison with 30 Day Delay ($\beta = 0.03$, $\beta' = 0.01$, $\bar{b} = 9$, $\mu = 0.35$, $\delta = 0.15$, $\delta' = 0.001$)

Given this explanation, we hypothesized that an “inverse proportional” allocation policy might be even more effective (smaller cities get proportionally more

²All data points are averaged over 100 independent runs.

vaccine). Our algorithm for computing inverse proportions is as follows [23]. Let N be the total population of the cities, n_i be the population of city i , and p_i be the proportion of the total population for city i , computed as $p_i = n_i/N$.

Let p_{max} be the proportion of the population of the largest city, $u_i = 1 - p_i$ (the proportion of the population not in city i), and m be the minimum of the values of u_i (the proportion of the population not in the largest city or $1 - p_{max}$). Then let $w_i = u_i/m$, and $W = \sum_{i=1}^{12} w_i$. Then $inv_i = w_i W$ are the inverse proportions of the city populations. In this model,

$$inv_i = \frac{1 - p_i}{1 - p_{max}} \sum_{i=1}^{12} \frac{1 - p_i}{1 - p_{max}}. \quad (15)$$

The algorithm is shown in Figure 6.

Name	Description
Inverse Proportional	Fully vaccinate the city of origin, distributing the rest to the remaining cities such that each city gets a vaccine supply inversely proportional to its proportion of the population (smaller cities get more than larger cities).

Figure 6: Inverse Proportional Baseline Vaccine Allocation Policy

Figure 7 shows how the new inverse proportional vaccine allocation policy compares with proportional and uniform vaccine allocation policies. The inverse proportional policy is competitive with or better than the prior two policies.

5. Evolved Vaccination Policies

Are there even better vaccination policies? We have a well-defined optimization problem, namely, to allocate vaccines such that the number of sick days is minimized. Hence, we turned to evolutionary algorithms (EAs) to automate our search for better vaccination policies. EAs have had excellent success at solving real-world problems such as in forensics [17], stock price forecasting [7], data compression in wireless sensor networks [22], path planning [35], and evolving static output feedback controllers [34]. If the EA could find better policies, our hope was that we could analyze the policies to explain what the EA had discovered.

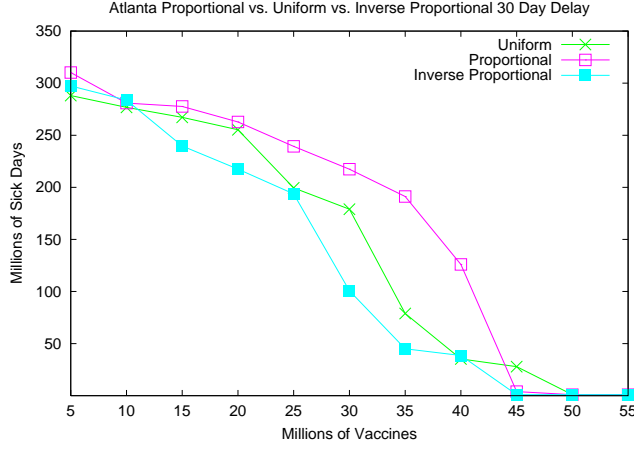


Figure 7: Atlanta Inverse Proportional Policy Baseline Comparison with 30 Day Delay

EAs are inspired by evolutionary theory, especially the concept of “survival of the fittest”. The EA maintains a population of individuals. Each individual represents a possible solution to a problem. A fitness function converts the representation into a real number that measures the quality of the solution. The EA takes a population of solutions and evolves them to a better fitness level. Selection serves to focus the search, and exploit the knowledge gained thus far. Genetic operators provide a random component, allowing for exploration of the search space. Patel, Longini and Halloran [24] also use an EA to evolve optimal vaccination strategies. However, their focus is on the optimal allocation of vaccine to five age groups: pre-school, school, young adults, middle aged adults, and old adults. Our focus is entirely different (although complementary), since we are concerned with the optimal allocation of vaccines to geographic regions of the country.

Specifically, we used a (μ, λ) Evolution Strategy (ES) with $\mu = 10$ parents creating $\lambda = 30$ children via mutation (described below). The $\mu = 10$ best children then constitute the population at the next generation. The basic execution of our ES is the following:

- 1) Randomly initialize the population of μ parents.
- 2) Evaluate the population using the fitness function.
- 3) Apply mutation to create λ children.
- 4) Evaluate the children using the fitness function.
- 5) Select the μ best children to form the next population.
- 6) Repeat 3-5 until some termination criteria are met.

In our case, an individual represents the allocation of vaccines to each of the twelve cities. We use an integer vector of length twelve, $\vec{v} = (v_1, v_2, \dots, v_{12})$, where v_i represents the number of vaccines allocated to city i . The values are constrained such that the sum of the vaccines allocated to each city equals the total number of available vaccines ($V = \sum_{i=1}^{12} v_i$).³ We created a “delta-swap” genetic operator to mutate the individuals. Two cities swap a random amount of vaccine. This preserves the constraint that $V = \sum_{i=1}^{12} v_i$. We also constrain the number of vaccines allocated to a city to be no greater than the population size of that city ($v_i \leq n_i$).

The details of “delta-swap” are as follows. Two cities are chosen uniformly randomly (however, the two cities can not be the same). Denote S as the source city (not to be confused with the Susceptible state of health), while D is the city of destination. Let v_S be the amount of vaccine currently at city S , while v_D is the amount of vaccine at city D . The amount of vaccine to swap, denoted Δv , is chosen uniformly from $U(0, \min(v_S, 200000))$. In other words, the amount of vaccine to swap is chosen uniformly between 0 and 200000, or 0 and v_S (if $v_S < 200000$). Finally, if $v_D + \Delta v < n_D$ the swap occurs and Δv vaccine is added to D and is subtracted from S .

All that remains is to measure the “fitness” of an individual. For this application, fitness is simply the number of sick days that occur, given the allocation of vaccines defined by the individual. Since we are minimizing, “better fitness” means that we want a lower number of sick days. One complication is that our viral simulation is stochastic. Hence, the fitness must be averaged over a number of independent trials. If the number is too large, the EA is too slow. If the number is too small, we can not trust the results (due to variance). We chose an alternative approach, based on our work where we evolve finite-state machines in non-deterministic environments [31, 32]. First the EA uses a relatively small number of trials (10) to estimate the fitness of an individual. If the fitness is better than the best fitness ever seen, then the EA re-evaluates that individual over a larger number of trials (40). This approach worked quite well. Our termination criterion was 100 generations, after which the best individual was evaluated over 100 runs of the simulation as with our benchmark policies.

Figure 8 shows the results for Atlanta with a 30 day delay in introducing the vaccine. The results are very impressive. The policy found by the EA is clearly

³Although our representation is fixed length, our motivation came from the variable length representation in the *proportional* genetic algorithm as described in [36].

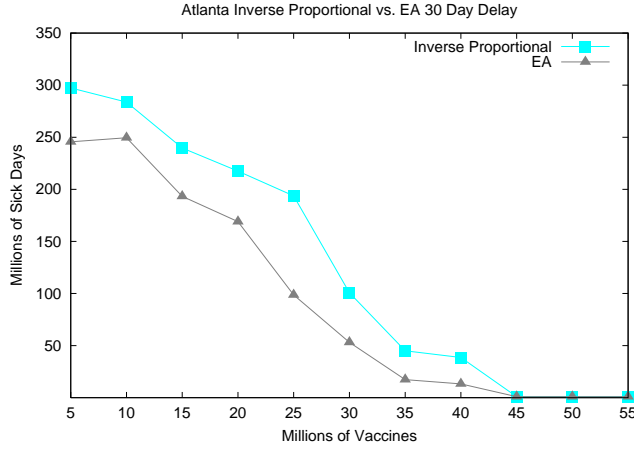


Figure 8: Atlanta EA and Inverse Proportional Policy Baseline Comparison with 30 Day Delay

superior to our inverse proportional policy.

6. A Better Vaccination Policy

Recall that that EA policies are simply vectors of twelve integers. As such, the reason for the good performance is not immediately obvious. Hence, we conducted a thorough analysis of a number of good EA policies, and found a number of commonalities.

First, as expected, the city of origin always received a large number of vaccines. Second, (and also expected), the EA gave more vaccine to smaller cities. This is in line with our inverse proportional policy. However, the EA policy is superior to the inverse proportional policy. Why is this the case? The key is to note what information the EA has implicitly available. The first commonality was based on figuring out the city of origin. The second commonality was based on the EA taking advantage of population size.⁴

However, there is a third source of important information that we have not taken into account, namely, the probability transition matrix Q . When we took this into consideration we found that the EA was not only focusing on small cities, but on cities that are flown to most often! The Q -matrix gives us the likely next

⁴It is important to clarify that the EA does not reason at this level. However, the EA is perfectly capable of taking advantage of useful features and structures in the search space. We are providing a human interpretation of these features and structures.

location for each plane, so it makes sense that this information is incredibly important in predicting where the virus spreads from a certain city or cities and thus contributes greatly to better vaccine allocation policies.

In order to mimic what the EA is doing, we created an “importance factor”. This factor captures the notion that the cities that are important to vaccinate are those that are small and are flown to most often. Let \vec{S}_0 be the initial distribution of the virus. For example, if the virus starts in Las Vegas, then our initial distribution is $\vec{S}_0 = [0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]$. If the virus starts in Atlanta and Las Vegas, then $\vec{S}_0 = [0.5, 0, 0, 0, 0, 0, 0.5, 0, 0, 0, 0, 0]$. Then $\vec{S}_1 = \vec{S}_0 \cdot Q$ represents the probability distribution of where the virus will travel, in the next time step. Let n_i be the population size of city i , and imp_i denote the importance of city i , $imp_i = S_{1,i}/n_i$.

Cities are then vaccinated in descending order based on their importance. However, is it necessary to fully vaccinate the important cities? We also noted that the EA tended to vaccinate roughly 92% of the city population! Recall, that for the experiments reported in this paper, theory indicated that at least 92% of the population of a city must be vaccinated, in order to rapidly eliminate the virus ($f_e \approx 92\%$). The EA clearly learned that providing more vaccine was wasteful, and that the excess should be distributed to the other cities.

Figure 9 describes a new vaccine allocation policy based on the city of origin, the predicted importance of a city, and the vaccine efficacy.

<i>Name</i>	Description
Markov	Fully vaccinate the city of origin, then vaccinate the rest in order of importance (those cities most flown to with smaller population sizes) up to the steady state level of vaccination f_e

Figure 9: Policy Combining Q -Matrix, Population Size, and Vaccine Efficacy

Figure 10 shows the results with the new policy. It is competitive with the EA policy, and is better at 15 million and 20 million vaccines. The only weakness is at 5 million vaccines. However, it is clear that we have captured and explained much of the EA policy.

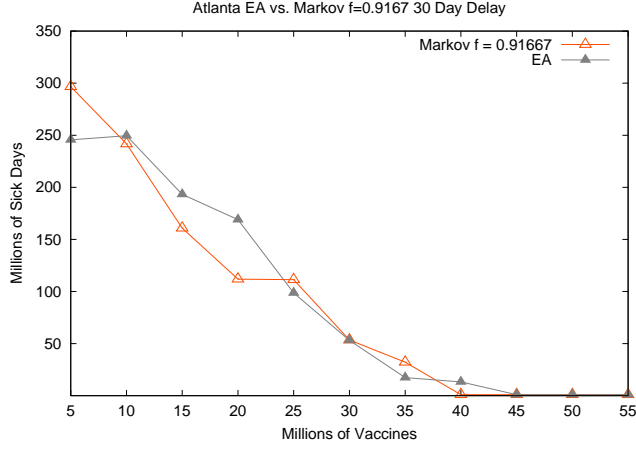


Figure 10: Atlanta EA and Markov Policy Baseline Comparison with 30 Day Delay

7. Confirmation of the Model

The key to our new vaccination strategy is the importance factor defined above. Although we cannot (currently) evaluate the vaccination strategy in the real world, we realized that we can in fact evaluate the importance factor itself. As stated above, the importance factor captures the central concept that the virus will have a greater impact on small cities that are flown to most often. If this importance factor is correct, it should be consistent with the spread of the real influenza throughout the U.S.

We examined the spread of the last four influenza seasons, using the CDC influenza maps. Our twelve cities are based in ten states - Arizona, California, Colorado, Florida, Georgia, Illinois, Michigan, Minnesota, Nevada, and Texas. We examined the situation where the influenza became “widespread” in a state (or states), and then we used the importance factor to predict where the influenza would spread to next. Widespread influenza activity is defined as “Outbreaks of influenza or increases in ILI [influenza-like illness] cases and recent laboratory-confirmed influenza in at least half the regions of the state.”[6]

7.1. 2005/2006 Influenza Season

In week 51 of 2005, the influenza virus reached widespread levels in California and Arizona (Figure 11). We split the infected percentages between Arizona and California, dividing it amongst the cities in those states. This gave us an initial distribution of $\vec{S}_0 = [0, 0, 0.3, 0, 0, 0, 0.3, 0, 0, 0, 0, 0.3, 0]$. We computed the importance

factors for each city (state). Three states had importance values much greater than the rest (Texas, Nevada, and Colorado). In week 52 of 2005 (Figure 12) and weeks 1 and 2 of 2006, these three states attain and maintain a widespread influenza activity level. This particular season was not especially virulent.

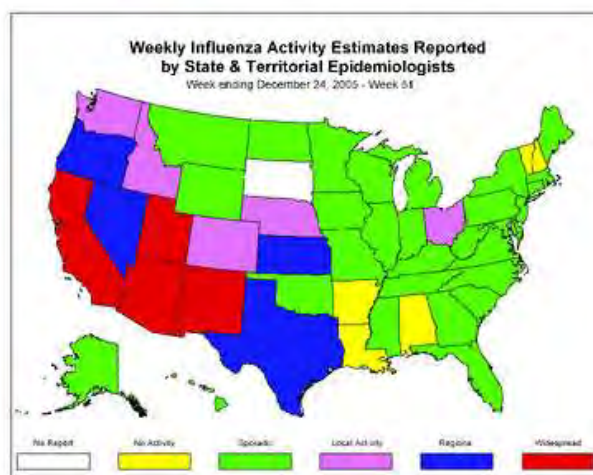


Figure 11: U.S. Influenza Activity Week Ending December 24, 2005

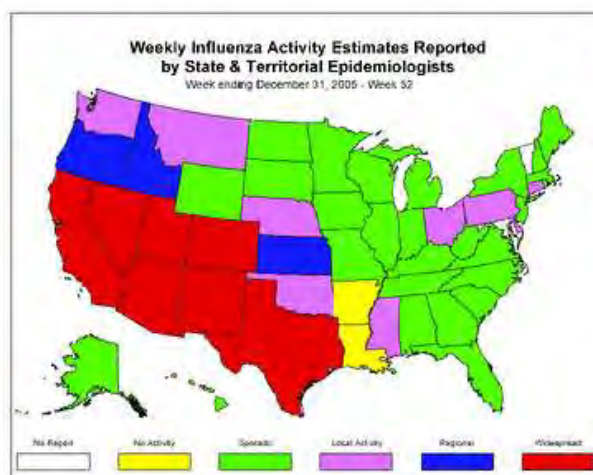


Figure 12: U.S. Influenza Activity Week Ending December 31, 2005

7.2. 2006/2007 Influenza Season

These 2005/2006 results are pleasing, but are a bit weak, since the states are adjacent. The 2006/2007 influenza season, however, was more interesting. In week 49 of 2006, influenza was widespread in Florida (Figure 13). If we use the initial distribution $\vec{S}_0 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]$, our model predicts that five states had importance factors much greater than the rest (in order, Georgia, and then to Colorado, Texas, Nevada, and Minnesota). This influenza is slower to spread, yet our predictions were surprisingly accurate.

In week 50 of 2006, we see that Georgia and Florida have widespread influenza conditions, which corresponds to the prediction that Atlanta would be next (Figure 14). We then have to jump ahead to week 4 of 2007, where Texas and Minnesota have become widespread (Figure 15). If we jump ahead another 4 weeks to week 8 of 2007, we see Colorado is now widespread (Figure 16). Our only important error was with respect to Nevada, which did not become widespread. However, it is important to note that these states are not adjacent, so it does seem reasonable to conclude that the spread was via air travel.

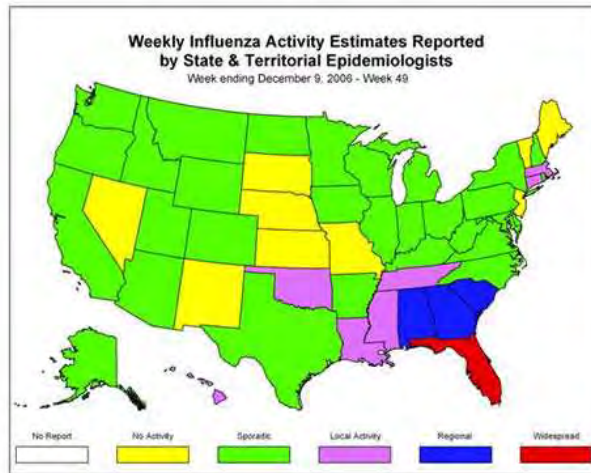


Figure 13: U.S. Influenza Activity Week Ending December 9, 2006

7.3. 2007/2008 Influenza Season

In week 49 of 2007, the influenza was widespread in Texas, so we started our initial distribution in Houston and Dallas. Our model predicts that the virus should spread first to Colorado, and then to Nevada, Georgia, Arizona, Minnesota,

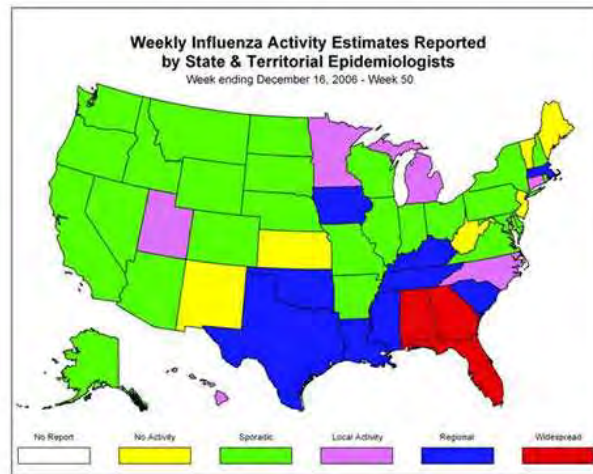


Figure 14: U.S. Influenza Activity Week Ending December 16, 2006

California, Michigan, Illinois, and Florida (in that order). This influenza was especially virulent – every state became widespread.

In week 1 of 2008, Colorado became widespread. Then, in week 5 of 2008 both Georgia and Arizona were widespread. In week 6 the influenza became widespread in Nevada, Minnesota, California, Michigan, and Illinois. Finally, in week 9 Florida became widespread. Our only error was again with respect to Nevada, which became widespread after Georgia and Arizona.

7.4. 2008/2009 Influenza Season

In week 1 of 2009 the influenza became widespread in Virginia. However, our model does not include any airports in Virginia, so we could not begin there. However, the second state to become widespread was Texas in week 3 of 2009. Hence, we ran the same model as with the 2007/2008 influenza season. This season was not as virulent – quite a few states never became widespread.

Of the states we model, Colorado became widespread next, in week 4. Then in week 5 both Nevada and Georgia became widespread. In week 6 Arizona and Florida became widespread. Finally, California became widespread in week 11. Minnesota, Michigan, and Illinois never became widespread. Our model has a minor error in that Minnesota never became widespread, although California did. However, an investigation of the importance factors showed that the numbers for both states were almost identical. Recall that the influenza really started in Virginia – this quite naturally has an affect on the real-world spread which

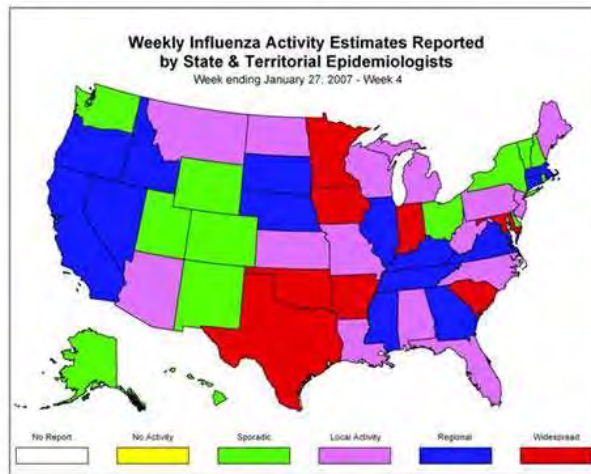


Figure 15: U.S. Influenza Activity Week Ending January 27, 2007

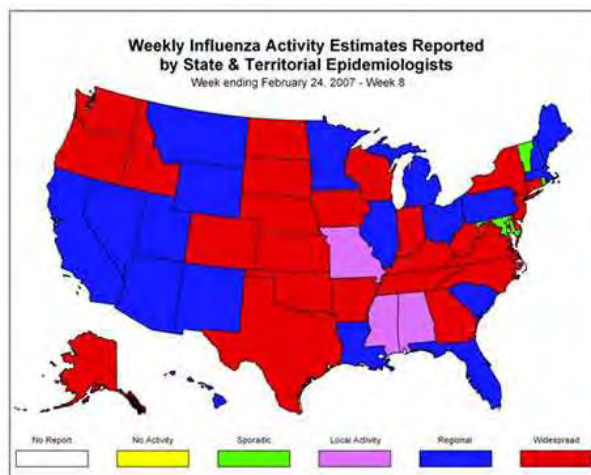


Figure 16: U.S. Influenza Activity Week Ending February 24, 2007

we could not model. This is reflected in our one significant error. Florida became widespread very early in the season, as opposed to very late. Again we consider this to be caused by Virginia (there are a lot of flights from the Virginia/Washington D.C./Maryland airports to Florida).

7.5. 2009/2010 Current Influenza Season

The CDC influenza maps monitor “Influenza-like Illnesses” (ILI), but don’t separate the different versions. Hence the surveillance reports do not or cannot subtype influenza A into H1, H3, and H1N1. Due to the longevity of H1N1, and the mixed data, it is impossible to separate out the normal influenza strains in order to make predictions.

7.6. Quantitative Confirmation of the Model

In this subsection we quantitatively analyze the accuracy of our predictions. As an illustrative example, consider the 2008/2009 influenza season. Colorado became widespread on January 31, both Nevada and Georgia became widespread on February 7, and both Arizona and Florida became widespread on February 14. Finally, California became widespread on March 21. Since the CDC maps are updated weekly, the result is a partial ordering of r states: $CO < NV \leq GA < AZ \leq FL < CA$ where $x < y$ means x occurs temporally before y (note, we adopt the USPS state abbreviations for conciseness). There are four possible strict total orderings resulting from the partial ordering:

1. $CO < NV < GA < AZ < FL < CA$
2. $CO < GA < NV < AZ < FL < CA$
3. $CO < NV < GA < FL < AZ < CA$
4. $CO < GA < NV < FL < AZ < CA$

As shown above, our model uses the importance factors to provide a strict total ordering of p states. Because not all states will necessarily become widespread, $r \leq p$. For the 2008/2009 influenza season our model predicts the following total ordering: $CO < NV < GA < AZ < MN < CA < MI < IL < FL$.

We now want to create a distance metric from the predicted ordering to the real world partial ordering. Our distance metric is the minimum number of “2 swaps” required to make the predicted ordering “consistent” with the real world partial ordering⁵. The predicted ordering is consistent if the first r cities of the predicted ordering match one of the real world total orderings.

In the case given above, if MN and FL are swapped in the predicted ordering, we get $CO < NV < GA < AZ < FL < CA < MI < IL < MN$. This is consistent with the real world partial ordering, so the distance of the predicted ordering from

⁵The 2 swap operator randomly chooses two different states and swaps their positions.

Influenza Season	Model δ	Random Predictions X			Statistics	
		μ_X	σ_X	$skew(X)$	$\text{Prob}(X \leq \delta)$	Wilcoxon p
2005/2006	2	3.34	0.85	-0.29	12%	< 0.00001
2006/2007	2	3.22	0.74	-0.72	14%	< 0.00001
2007/2008	1	3.17	0.75	-0.56	2%	< 0.00001
2008/2009	1	4.23	1.02	-0.80	1%	< 0.00001

Table 4: Quantitative Analysis

reality is one. For the first three influenza seasons (2005/2006, 2006/2007, and 2007/2008) the distance from reality is two, two, and one, respectively. Hence, all predictions were close to reality, with the last two seasons having the best predictions. The “ δ ” column in Table 4 summarizes those results.

As a baseline comparison, we also created 100 random predictions for each of the four influenza seasons. If our model is correctly capturing the dynamics of the real world, the predictions from the model should be significantly better than predictions provided by random orderings. The random variable X denotes the distance of the random prediction from the real world. The mean, standard deviation and skewness of X are also shown in Table 4, denoted by μ_X , σ_X , and $skew(X)$. In all cases the mean distance of the random distributions is higher than the prediction created by our model. For the first two influenza seasons the quality of model prediction is roughly 1.5 standard deviations better than the mean. For the last two influenza seasons the quality of the model prediction is roughly three standard deviations better than the mean.

It is helpful to calculate the probability that the random predictions could match or outperform the quality of the model prediction. Those probabilities are given in Table 4, under $\text{Prob}(X \leq \delta)$. For all four seasons, the probabilities are 12%, 14%, 2% and 1%. Treated as a whole, it is therefore highly unlikely for random predictions to perform nearly as well as our model.

Although all results indicate that our model is performing better than random, the results for the last two seasons are exceptionally good. We believe the reason for this is that we only created one Q matrix, reflecting current flight usage. Flight usage varies by year, and it is quite reasonable that the current Q matrix would not model older seasons as accurately.

For one final test, we performed a Wilcoxon signed-rank test. This test does not assume any particular underlying distribution, and can be used when the distribution is reasonably symmetric. The values of $skew(X)$ (which are low) indicate that the Wilcoxon test can be applied. For all four seasons, the significance level p is much less than 0.00001, confirming again that it is extremely unlikely for the

quality of random predictions to be nearly as good as the quality of the model predictions.

7.7. Summary

Although we do not attempt to predict the speed of the viral spread (as explained below), the predicted ordering of viral spread is surprisingly accurate, given that we are modeling flights from only 12 cities.

8. Conclusion

We have created a framework for simulating and analyzing weakly connected island models. This framework consists of two levels. First, a micro-level describes the behavior of individuals within an island. Second, the macro-level models the travel of individuals between islands. The two levels are largely separate, from a functional point of view. The micro-level is used to calculate the *speed* of the viral spread. The macro-level is used to calculate *where* the virus will spread. The key to the accurate modeling of the micro-level is to fit the virus parameters to the observable quantities of a particular virus. However, once this is done, it is extremely difficult to draw general conclusions. This is the reason why we modeled a rather generic influenza virus and a small number of airports - it lends itself to a reductionist interpretation far more readily.

Our reductionist approach serves as a contrast to the work with large epidemiological “engines”, where the models are extremely difficult to analyze. One example is the work by Colizza, et al. [9], that uses the largest 3,100 airports.⁶ However, the model is then used to simulate the spread of only one influenza virus in only nine “surveillance regions” within the U.S., as opposed to examining individual states. No reductionist interpretation is provided.

We applied our model to the spread of a virus between cities via air travel. This model has been verified and validated, by demonstrating that the steady state of the micro-level (differential equations) and the macro-level (Markov) agree with mathematical analysis. We then augmented the model to include vaccinations and used an EA optimizer to generate good policies. We then examined those policies to explain and develop superior policies. It is important to point out that the EA was used both as an optimizer and a tool for discovery. Without the EA it would have been extremely difficult for us to understand the dynamics of the model.

⁶International Air Transport Association database (<http://www.iata.org>)

The key to our success is an “importance factor”, which states that the cities of importance (both from the point of view of viral spread and their need for vaccines) are those that are small yet are flown to most often. Smaller cities experience a more rapid response to an invading infected population. They also produce infected individuals more quickly as a result. Cities that are flown to more often are dangerous because more planes are carrying infected people to these cities. This factor provides a highly valuable reductionist interpretation of a complex model.

We have also demonstrated that there is a connection between our importance factor and the spread of influenza in the real world by comparing the previous four influenza seasons to our importance factor predictions. This suggests that we have captured a good first-order approximation of the effect of air-travel as a virus vector on the United States. It is important to note that this success comes despite the fact that each influenza season involves a different virus with different micro-level parameters that we did not attempt to model. In fact, an alternative micro-level model omits the Asymptomatic state of health - namely, the SIR model. In this situation Infected people can fly. Our studies indicate that, once again, due to the functional separation between the micro-level and macro-level models, this change affects the speed of the virus (as would be expected), but not the ordering.

Due to the number of studies we have performed (there is insufficient room in this paper to include them all), we believe that these observations are quite general and hold in a wide variety of circumstances [29]. However, the nice aspect of our framework is that if conditions change radically, the EA can always be invoked by the user to generate new policies automatically.

In this paper we developed the “importance factor” via human inspection of the data (a set of high fitness EA individuals). However, suppose we include additional cities from other countries. Then there could be significant changes in \bar{b} (due to differences in population density in different countries). Although our simulation framework would have no difficulty with these changes, and the EA could still be applied, human inspection of the resulting high fitness individuals will become much more difficult. One of the main challenges is that our human reasoning involved more than just inspection of the individuals, but required an understanding of the underlying simulation model, including the city of origin, the size of the populations, the proper interpretation of the probability transition matrix, and the vaccine feasibility region. Hence, future work will require the addition of sophisticated data mining (e.g., [19, 15, 14]) and equation discovery components [5, 33, 3], to automate the reductionist approach taken in this paper.

Acknowledgment

Lora Billings was supported by ARO grant W911NF-06-1-0320. Lucas Shaw was supported by the National Institutes of Health.

References

- [1] Allen, L. J. S., 1994. Some discrete-time SI, SIR, and SIS epidemic models. *Mathematical Biosciences* 124, 83–105.
- [2] Arino, J., Jordan, R., van den Driessche, P., 2007. Quarantine in a multi-species epidemic model with spatial dynamics. *Mathematical Biosciences* 206, 46–60.
- [3] Bay, S. D., Shapiro, D. G., Langley, P., 2003. Revising engineering models: Combining computational discovery with knowledge. In: *Proceedings of the Thirteenth European Conference on Machine Learning*.
- [4] Billings, L., Spears, W. M., Schwartz, I. B., 2002. A unified prediction of computer virus spread in connected networks. *Physics Letters A* 297, 261–266.
- [5] Bridewell, W., Langley, P., Todorovski, L., Dzeroski, S., 2008. Inductive process modeling. *Machine Learning* 71, 1–32.
- [6] Centers for Disease Control (CDC), 2006. CDC - Influenza (Flu) | Flu Activity. Online; <http://www.cdc.gov/flu/weekly/fluactivity.htm>.
- [7] Cheng, C. H., Chen, T. L., Wie, L. Y., 2010. A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Information Sciences* 180, 1610–1629.
- [8] Colizza, V., Barrat, A., Barthelemy, M., Vespignani, A., 2006. The modeling of global epidemics: Stochastic dynamics and predictability. *Bulletin of Mathematical Biology* 68, 1893–1921.
- [9] Colizza, V., Barrat, A., Barthelemy, M., Vespignani, A., 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences* 103 (7), 2015–2020.

- [10] Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., Burke, D. S., 2006. Strategies for mitigating an influenza pandemic. *Nature* 442, 448–452.
- [11] Grais, R. F., Ellis, J. H., Glass, G. E., OCT 2003. Forecasting the geographical spread of smallpox cases by air travel. *Epidemiology and Infection* 131 (2), 849–857.
- [12] Hethcote, H. W., Aug. 2005. The Basic Epidemiology Models I & II: Models, Expression for R_0 , Parameter Estimation, and Applications. Lecture Note Series, Workshop in Mathematical Modeling of Infectious Diseases: Dynamics and Control. Institute for Mathematical Sciences, National University of Singapore.
- [13] Hethcote, H. W., Aug. 2005. Epidemiology Models with Variable Population Size. Lecture Note Series, Workshop in Mathematical Modeling of Infectious Diseases: Dynamics and Control. Institute for Mathematical Sciences, National University of Singapore.
- [14] Hong, C. F., 2009. Qualitative chance discovery extracting competitive advantages. *Information Sciences* 179, 1570–1583.
- [15] Hu, H L, C. Y. L., 2008. Mining typical patterns from databases. *Information Sciences* 178, 3683–3696.
- [16] Hyman, J. M., LaForce, T., 2003. Modeling the spread of influenza among cities. In: Banks, H. T., Castillo-Chavez, C. (Eds.), *Bioterrorism: Mathematical Modeling Applications in Homeland Security*. SIAM, p. 211.
- [17] Ibáñez, O., Ballerini, L., Córdón, O., Damas, S., Santamaría, J., 2009. An experimental study on the applicability of evolutionary algorithms to cranio-facial superimposition in forensic identification. *Information Sciences* 179, 3998–4028.
- [18] Liebovitch, L. S., Schwartz, I. B., 2004. Migration induced epidemics: dynamics of flux-based multipatch models. *Physics Letters A* 332, 256–267.
- [19] Liu, H., Wang, X., He, J., Xin, D., Shao, Z., 2009. Top-down mining of frequent closed patterns from very high dimensional data. *Information Sciences* 179, 899–924.

- [20] Lloyd, A. L., May, R. M., 1996. Spatial heterogeneity in epidemic models. *J. Theoretical Biology* 179, 1–11.
- [21] Longini, I. M., Halloran, M. E., Nizam, A., Yang, Y., Xu, S. F., Burke, D. S., Cummings, D. A. T., Epstein, J. M., 2007. Containing a large bioterrorist smallpox attack: a computer simulation approach. *International Journal of Infectious Diseases* 11, 98–108.
- [22] Marcelloni, F., Vecchio, M., 2010. Enabling energy-efficient and lossy-aware data compression in wireless sensor networks by multi-objective evolutionary optimization. *Information Sciences* 180, 1924–1941.
- [23] Morrison, R. W., 2006. E-mail correspondence on inverse proportional selection. *E-mail Correspondence*.
- [24] Patel, R., Longini, I. M., Halloran, M. E., 2005. Finding optimal vaccination strategies for pandemic influenza using genetic algorithms. *Journal of Theoretical Biology* 234, 201–212.
- [25] Piccolo, C., Billings, L., 2005. The effect of vaccinations in an immigrant model. *Mathematical And Computer Modelling* 42, 291–299.
- [26] Rvachev, L. A., Longini, I. M., 1985. A mathematical model for the global spread of influenza. *Mathematical Biosciences* 75, 3–22.
- [27] Shaw, L., 2006. A Parallel Evolutionary Algorithms Framework. Technical report, Department of Computer Science, University of Wyoming.
- [28] Shaw, L., 2006. PEA User’s Guide. User’s manual, Department of Computer Science, University of Wyoming.
- [29] Shaw, L. R., 2007. A Computational Framework For Modeling The Spread Of Pathogens And Generating Effective Containment Strategies In Weakly Connected Island Models. Masters thesis, University of Wyoming.
- [30] Spears, W. M., Billings, L., Schwartz, I. B., Feb. 2001. Modeling Viral Epidemiology in Connected Networks. Memorandum Report NRL/MR/6700–01-8537, Naval Research Laboratory (NRL).
- [31] Spears, W. M., Gordon[-Spears], D. F., 2000. Evolving finite-state machine strategies for protecting resources. In: *Proceedings of ISMIS’00*.

- [32] Spears, W. M., Gordon-Spears, D. F., 2003. Evolution of strategies for resource protection problems. In: *Advances in evolutionary computing: theory and applications*. Springer-Verlag New York, Inc., New York, NY, USA, pp. 367–392.
- [33] Todorovski, L., Dzeroski, S., Langley, P., Potter, C., 2003. Using equation discovery to revise an earth ecosystem model of carbon net production. *Ecological Modelling* 170, 141–154.
- [34] Toscano, R., Lyonnet, P., 2010. Robust static output feedback controller synthesis using Kharitonov’s theorem and evolutionary algorithms. *Information Sciences* 180, 2023–2028.
- [35] UĞUR, A., 2008. Path planning on a cuboid using genetic algorithms. *Information Sciences* 178, 3275–3287.
- [36] Wu, A. S., Garibay, I., 2002. The proportional genetic algorithm: Gene expression in a genetic algorithm. *Genetic Programming and Evolvable Machines* 3 (2), 157–192.